Title

The evolution of tenascins and fibronectin.

Journal

Cell adhesion & migration, 9(1-2)

ISSN

1933-6918

Authors

Adams, Josephine C
Chiquet-Ehrismann, Ruth
Tucker, Richard P

Publication Date

2015

DOI

10.4161/19336918.2014.970030

Peer reviewed

Cell Adhesion & Migration 9:1-2, 22–33; January–April 2015; Published with License by Taylor & Francis Group, LLC

The evolution of tenascins and fibronectin

Josephine C Adams[1], Ruth Chiquet-Ehrismann[2,3], and Richard P Tucker[4,*]

[1]School of Biochemistry, University of Bristol; Bristol, UK; [2]Friedrich Miescher Institute for Biomedical Research; Basel, Switzerland; [3]University of Basel; Faculty of Science; Basel, Switzerland; [4]Department of Cell Biology and Human Anatomy, University of California at Davis; Davis, CA USA

Keywords: coelacanth, elephant shark, extracellular matrix, integrin, lamprey, phylogenomics

Tenascins are extracellular matrix glycoproteins that act both as integrin ligands and as modifiers of fibronectin-integrin interactions to regulate cell adhesion, migration, proliferation and differentiation. In tetrapods, both tenascins and fibronectin bind to integrins via RGD and LDV-type tripeptide motifs found in exposed loops in their fibronectin-type III domains. We previously showed that tenascins appeared early in the chordate lineage and are represented by single genes in extant cephalochordates and tunicates. Here we have examined the genomes of the coelacanth *Latimeria chalumnae*, the elephant shark *Callorhinchus milii* as well as the lampreys *Petromyzon marinus* and *Lethenteron japonicum* to learn more about the evolution of the tenascin gene family as well as the timing of the appearance of fibronectin during chordate evolution. The coelacanth has 4 tenascins that are more similar to tetrapod tenascins than are tenascins from ray-finned fishes. In contrast, only 2 tenascins were identified in the elephant shark and the Japanese lamprey *L. japonicum*. An RGD motif exposed to integrin binding is observed in tenascins from many, but not all, classes of chordates. Tetrapods that lack this RGD motif in tenascin-C have a similar motif in the paralog tenascin-W, suggesting the potential for some overlapping function. A predicted fibronectin with the same domain organization as the fibronectin from tetrapods is found in the sea lamprey *P. marinus* but not in tunicates, leading us to infer that fibronectin first appeared in vertebrates. The motifs that recognize LDV-type integrin receptors are conserved in fibronectins from a broad spectrum of vertebrates, but the RGD integrin-binding motif may have evolved in gnathostomes.

Introduction

Tenascins are a family of large extracellular matrix glycoproteins composed of repeated epidermal growth factor (EGF)-like domains, fibronectin-type III (FNIII) domains and a C-terminal fibrinogen related domain (FReD). Near the N-terminus is a region that forms a coiled-coil, and most tenascins are believed to exist as trimers or as hexamers formed from 2 trimers held together with disulfide bonds. Tenascins have several identified roles in cell adhesion and migration during development, tissue homeostasis and responses to disease or trauma, many of which are related to their ability to signal directly through integrin receptors or by binding to the extracellular matrix glycoprotein fibronectin and influencing the way that fibronectin signals through integrins (for review, see Chiquet-Ehrismann and Tucker[1]). Both tenascins and fibronectins can bind to integrins via the tripeptide motif arginine-glycine-aspartic acid (RGD), which is characteristically present in an exposed loop between the F and G β strands (the "F-G loop") of specific FNIII domains.[2]

We have shown that tenascins are relative newcomers to the extracellular matrix that first appeared early in chordate evolution, as evidenced by the presence of single tenascin genes in the cephalochordate *Branchiostoma floridae* and the tunicates *Ciona intestinalis* and *C. savignyi*.[3] The RGD motif is missing from these tunicate tenascins but is present in the F-G loop of many of the FNIII domains of the cephalochordate tenascin. As genomic evidence supports tunicates being the sister group to the vertebrates,[4] this led us to conclude that integrin binding via an RGD motif was a characteristic of the first tenascins that has been lost in the *Ciona* lineage of tunicates.

Fibronectin has a highly conserved domain architecture composed of fibronectin-type I (FNI), fibronectin-type II (FNII) and FNIII domains.[5] A fibronectin-like predicted protein is found in tunicates, but it contains immunoglobulin-like domains and lacks the stereotypical number and order of FNI, FNII and FNIII domains found in the true fibronectin of vertebrates.[3]

In tetrapods (amphibians, reptiles, birds and mammals) there are 4 tenascin genes: tenascin-C, tenascin-R, tenascin-W and tenascin-X (**Fig. 1A**).[6] Tenascin-C is the original "tenascin" that was first described independently in the myotendinous junctions and brains of the chicken *Gallus gallus*[7,8] as well as in human glioblastoma-derived cells.[9] The best studied of the tenascins, tenascin-C is expressed by motile cells such as the neural crest, at sites of connective tissue differentiation and in the adult in numerous stem cell niches; it also reappears during inflammation, at the margins of healing wounds and in the stroma of solid tumors.[1,10,11] Tenascin-C from *G. gallus* binds to α9β1 integrins via the isoleucine-aspartic acid-glycine (IDG) motif found in the B-C loop of its third FNIII domain[12] and to αVβ3 integrins via an RGD motif in the F-G loop of the same FNIII domain.[13] Tenascin-C can also bind to fibronectin and prevent syndecan-4-fibronectin interactions, which in turn influences fibronectin signaling through integrins.[14,15,16] Tenascin-C knockout mice have

*Correspondence to: Richard P Tucker; Email: rptucker@ucdavis.edu
Submitted: 06/23/2014; Revised: 08/14/2014; Accepted: 08/25/2014
http://dx.doi.org/10.4161/19336918.2014.970030

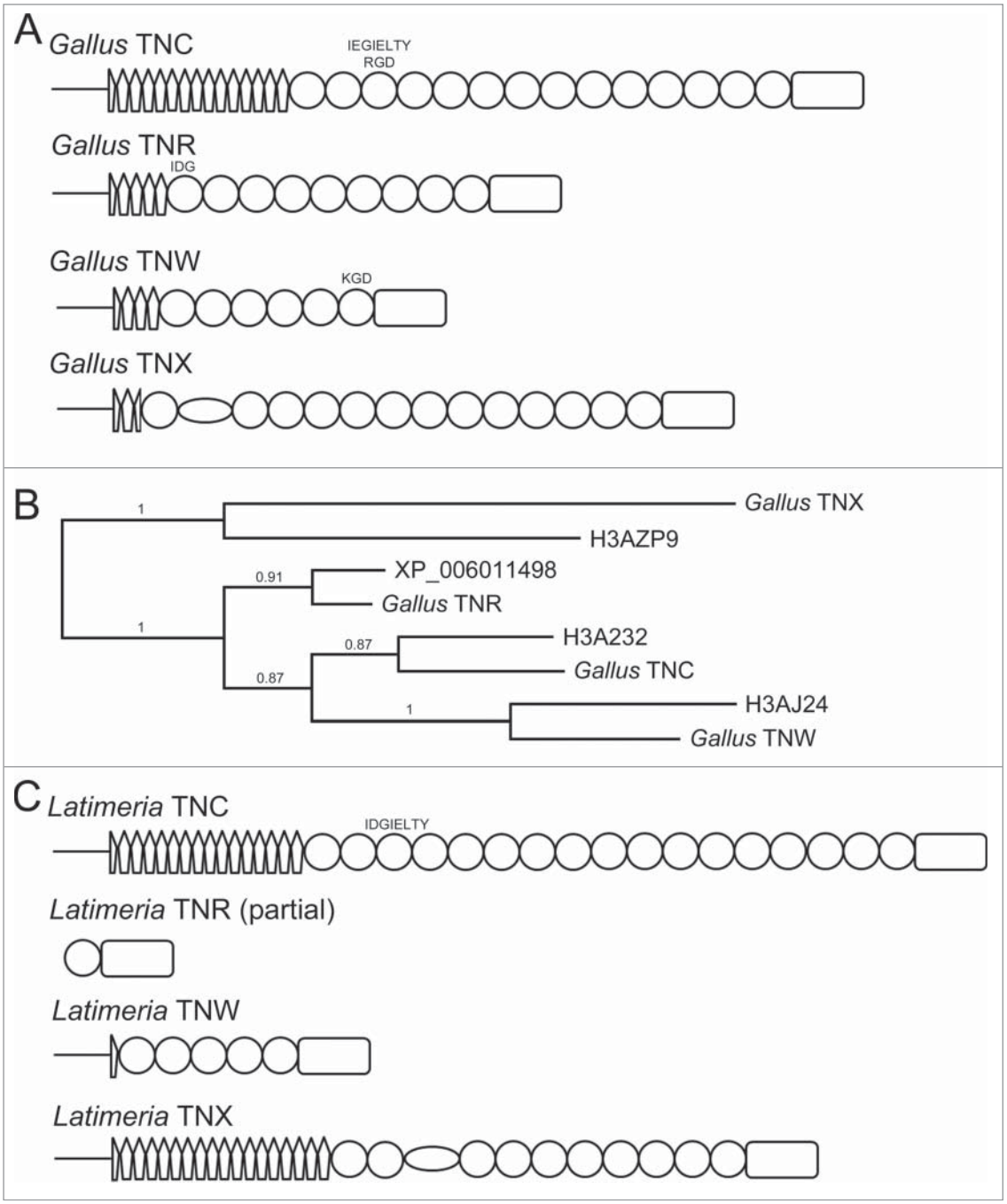

Figure 1. Four tenascin are found in tetrapods. (**A**) Those of the chicken *Gallus gallus* are illustrated schematically here. Epidermal growth factor-like domains are represented by the narrow pentagons, fibronectin-type III (FNIII) domains are represented by circles and the C-terminal fibrinogen-related domain is represented by the rectangle with rounded corners. A domain of unknown function that is unique to tenascin-X (TNX) is represented by an oval. Integrin recognition motifs found in FNIII domains are indicated. TNC, tenascin-C; TNR, tenascin-R; TNW, tenascin-W. (**B**) Four predicted tenascins are found in the coelacanth *Latimeria chalumnae*. They are identified here by their UniProt ID or by their GenBank accession number. Rooted phylogenetic tree analysis reveals their relationships to the 4 tenascins of *G. gallus*. Branch support is indicated. (**C**) The domain architecture of the *L. chalumnae* tenascins demonstrates their structural similarity to the homologs from tetrapods. The predicted TNR sequence is partial.

profound behavioral abnormalities and other phenotypes that are primarily associated with aberrant stem cell migration, proliferation and differentiation.[1,11] Tenascin-R, the second tenascin to be discovered, is restricted in its expression to the nervous system[17] and tenascin-R knockout mice also display abnormal behavior.[1] Tenascin-W is often co-expressed with tenascin-C both in development and in stem cell niches.[18] Murine tenascin-W has an RGD motif in the F-G loop of the second FNIII domain that appears to act as an $\alpha 8\beta 1$ integrin-binding site.[19] It and the tenascin-W of other species have the related lysine-

glycine-aspartic acid (KGD) motif in the F-G loop of one or more FNIII domains, and synthetic peptides that contain this motif and neighboring conserved sequences promote cell proliferation.[20] Tenascin-W knockout mice have not been analyzed. Finally, tenascin-X is widely expressed in loose connective tissue both in the adult and during development.[21] RGD and KGD motifs are notably absent from the F-G loops of the FNIII domains of tenascin-X. However, cell attachment to the FReD of tenascin-X is blocked with antibodies to β1 integrins.[22] Knockout of the tenascin-X gene in the mouse results in Ehlers-Danlos Syndrome (i.e., loose skin and hypermobile joints).[23]

It has been proposed that the diversity of the teleost fishes is due in part to the selective retention of genes following 3 rounds of whole genome duplication events.[24,25] Accordingly, there are 5 tenascins in the zebrafish *Danio rerio* and the green pufferfish *Tetraodon nigroviridis* instead of 4, with the additional tenascin resulting from a duplication of tenascin-C.[6] Some of the tenascin-C orthologs in the actinopterygians have exposed RGD motifs, but others do not. In *D. rerio* there are also 2 copies of the fibronectin gene, both of which are expressed.[26] RGD motifs are found in the F-G loop of the tenth FNIII domain of the fibronectins from ray-finned fishes, precisely where they are found in the fibronectins from tetrapods. Functional analyses of the RGD motifs in fish tenascin have not been reported, but RGD peptides block adhesion to zebrafish fibronectin *in vitro*.[27]

Recently, the genomic sequences from fish that are key to our understanding of chordate evolution were reported. These include the genomes of 2 jawless vertebrates of the lamprey family, the sea lamprey *Petromyzon marinus*[28] and the Japanese lamprey *Lethenteron japonicum*;[29] the cartilaginous elephant shark *Callorhinchus milii*;[30] and a lobe-finned fish, the coelacanth *Latimeria chalumnae*.[31] Almost nothing is known about tenascins in these species apart from detection of tenascin-C immunoreactivity in the lamprey olfactory pathway.[32] Here, we have identified and analyzed the genome-predicted tenascins and fibronectins from these and other species to learn more about the evolution of the tenascin gene family, the earliest appearance of true fibronectins, and the presence of RGD and LDV-type motifs as predicted integrin recognition sequences in tenascins and fibronectins during chordate evolution.

Results

The 4 tenascins of the coelacanth *Latimeria chalumnae*

Three predicted proteins with domain organizations characteristic of tenascins were identified in the genome of the coelacanth *L. chalumnae* using the graphical domain architecture feature of Pfam and searching for proteins with EGF-like domains and FNIII domains: H3A232, H3AZP9 and H3AJ24 (UniProt ID). A Delta Blast search with the H3A232 sequence against *L. chalumnae* sequences at NCBI revealed 3 isoforms (XP_006005618.1; XP_006005617.1; XP_006005616.1) that differ in the number of predicted FNIII domains. The largest (XP_006005616.1; "tenascin isotype X1") corresponds to H3A232 with additional N-terminal sequences. A fourth putative tenascin (XP_006011498.1) was identified by a tblastn search of *L. chalumnae* sequences with the C-terminal FReD sequence of the pufferfish *Takifugu rubripes* (XP_003975627). The relationships between the 4 tenascins from *L. chalumnae* and those of a representative tetrapod, the chicken *G. gallus*, were explored by construction of a rooted phylogenetic tree. H3A232/XP_006005616 clusters with tenascin-C (GenBank NP_990787 combined with variable domains CAA52055; **Fig. S1**), H3AZP9 with tenascin-X (CAA67509))), H3AJ24 with tenascin-W (AM231718), and XP_006011498 with tenascin-R (CAA45920) (**Fig. 1B**; **Fig. S2**).

L. chalumnae tenascin-C is composed of 15.5 EGF-like domains, 17 FNIII domains and a C-terminal FReD (**Fig. 1C**). The program LOGICOIL[33] predicts that the sequences corresponding to amino acids 117-148 form a coiled-coil, and that the most probable oligomer state is a trimer (**Table 1**). The third FNIII domain of tenascin-C from many tetrapods contains 2 experimentally-identified integrin binding motifs: the α9β1 integrin binding motif IDG[12] and the classic αVβ3-binding tripeptide motif RGD in the exposed F-G loop.[33] Alignment of the third FNIII domain from *G. gallus* and *L. chalumnae* tenascin-Cs reveals conservation of the α9β1-binding motif and neighboring sequences, but RGG is found in the F-G loop of the coelacanth tenascin instead of RGD (**Fig. 1C**). The predicted *L. chalumnae* tenascin-W is unusual in that it does not contain a complete EGF-like domain; only a partial EGF-like domain is predicted between the N-terminus and the first FNIII domain (**Fig. 1C**). The tripeptide motif KGD is found in the F-G loop of one or more FNIII domains in all tenascin-W sequences identified from tetrapods,[20] but the *L. chalumnae* tenascin-W lacks this motif. A coiled-coil motif was not identified, suggesting that the predicted sequence is incomplete. *L. chalumnae* tenascin-X is predicted to have 17.5 EGF-like domains, 10 FNIII domains and a FReD (**Fig. 1C**). Like the coelacanth tenascin-C it is predicted to assemble as a trimer (**Table 1**). Between the second and third FNIII domain is sequence that aligns with a similar region found in tetrapod tenascin-X. This is in contrast to the unrelated interrupting region found in tenascin-X from ray-finned fishes, which contains numerous copies of the 16 amino acid sequence GKEQKKATEGENTLSP.[6] The partial tenascin-R sequence (**Fig. 1C**) was not analyzed further.

Table 1. LOGICOIL oligomer state prediction of novel tenascins

Species/tenascin	Anti-parallel Dimer	Parallel Dimer	Trimer	Higher Order
Latimeria chalumnae Tenascin-C	0.95	1.01	**1.53**	0.82
Latimeria chalumnae Tenascin-X	0.92	1.10	**1.32**	1.04
Callorhinchus milii Tenascin-C (scaffold 45)	0.93	1.03	**1.53**	0.88
Callorhinchus milii Tenascin-R (scaffold 101)	0.93	1.04	**1.31**	1.14
Lethenteron japonicum Tenascin (scaffold 10)	1.00	0.94	**1.17**	1.04

The elephant shark *Callorhinchus milii* has 2 tenascins

There are 2 predicted tenascins in the genome of the cartilaginous holocephalan *Callorhinchus milii*. The first is found on scaffold 45 and is composed of 9.5 EGF-like domains, 18 FNIII domains and a C-terminal FReD. The second is found on scaffold 101 and is composed of 4.5 EGF-like domains, 10 FNIII domains and a C-terminal FReD (**Fig. 2**; **Fig. S3**). Both are predicted to form trimers via coiled-coils near their N-termini (**Table 1**). Blast searches with the C-terminal FReD of the scaffold 45 tenascin revealed significant sequence identity with tenascin-C. When the third FNIII domain of scaffold 45 tenascin is aligned with the same domain from human tenascin-C, the IDG-based α9β1 integrin binding region is well conserved, but the sequence RGQ is found in the F-G loop, not RGD (**Fig. 2**). Blast searches indicate that scaffold 101 tenascin is most similar to tenascin-R from other fish and tetrapods; like all other tenascin-R sequences examined it has an IDG in the first FNIII domain, but as yet there is no experimental evidence that tenascin-R is a ligand for integrins containing the α4 or α9 subunits (the integrin α subunits involved in binding to this motif in other extracellular matrix ligands).[35]

Tenascins from lampreys

A single, partial predicted tenascin is encoded in the genomic sequence of the sea lamprey *P. marinus* on scaffold GL480649 (ENSPMAT00000005710). The predicted protein has 29 FNIII domains and a C-terminal FReD (**Fig. 2**). The amino acid sequences of many of the FNIII domains are nearly identical, as has previously been noted in the tenascin from the cephalochordate *B. floridae*.[3] However, while the tenascin from *B. floridae* has multiple copies of RGD motifs in the F-G loops of its FNIII domains, none are found in the partial tenascin from this lamprey. There is, however, an IDG motif in the G-A loop between the fifth and fourth FNIII domains N-terminal to the FReD. In contrast, 2 predicted tenascins are found in the genomic sequences of the Japanese lamprey *Lethenteron japonicum*. One is found on scaffold 10 (KE993681). This tenascin has 3.5 EGF-like domains, 6 FN III domains and a C-terminal FReD (**Fig. 2**). An IDG motif is found in the G-A loop between the second and third FNIII domains and an RGD motif is found in the F-G loop of the fourth FNIII domain, suggesting that this tenascin is a potential ligand for multiple integrins. The second tenascin is found on scaffold 62 (KE993733) and has 6.5 EGF-like domains, 16 FNIII domains and a C-terminal FReD (**Fig. 2**). Like the partial tenascin from *P. marinus*, the scaffold 62 tenascin from the Japanese lamprey has an IDG motif in the linker region between 2 FNIII domains near the FReD. The scaffold 62 tenascin may have many more FNIII domains than are illustrated in **Figure 2**: the FNIII domains of this protein align precisely with a subset of FNIII domains from a large predicted protein (JL5413) composed of 43 serially arranged FNIII domains.

Phylogenetic analysis of tenascins

The relationships between representative tenascins and the novel tenascin sequences described above were analyzed by the construction of phylogenetic trees based on FReD amino acid sequences (**Fig. 3A**). The FReD of the *P. marinus* tenascin was not used for this analysis since alignment revealed it was likely to be incomplete. Moreover, preliminary phylogenetic trees based on all available amino acid sequence strongly indicated that the *P. marinus* tenascin is the homolog of the scaffold 62 tenascin from *L. japonicum* (results not shown). The phylogenetic tree confirms that the 4 tenascins from *Latimeria chalumnae* are homologs of tenascins-C, -R, -X and -W. The 2 tenascins

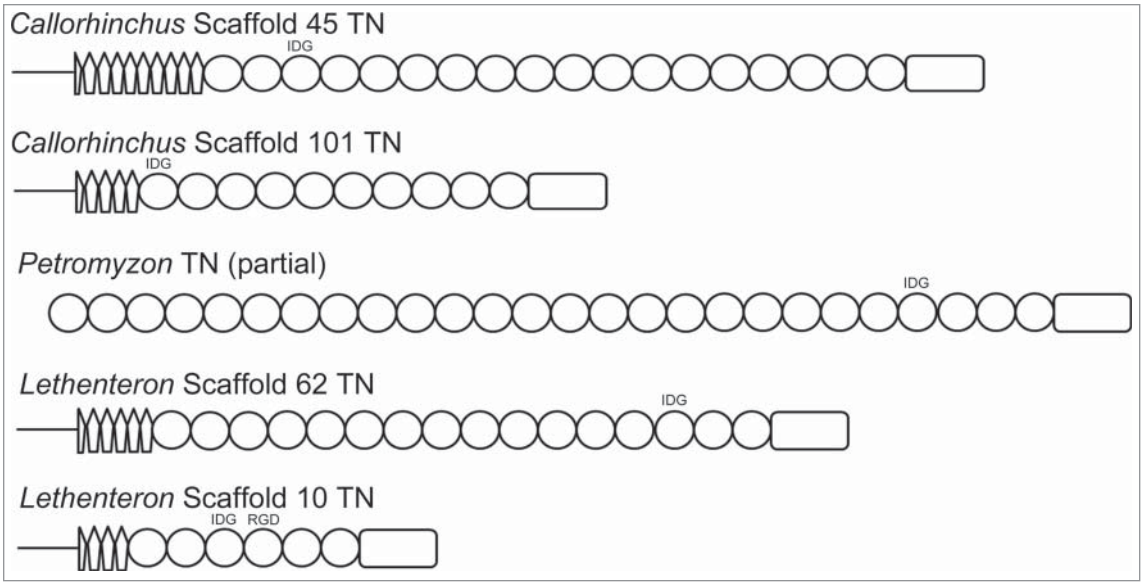

Figure 2. Two tenascins are found in the elephant shark *Callorhinchus milii*, one in the sea lamprey *Petromyzon marinus*, and 2 in the Japanese lamprey *Lethenteron japonicum*. The domain architecture is illustrated. See the legend from **Figure 1** for a key to the domains. Potential integrin-binding motifs present in exposed loops in FNIII domains are indicated.

identified in the genomic sequences of *C. milii* are located near the stem of the tenascin-R (scaffold 101 tenascin) and tenascin-C (scaffold 45 tenascin) clades, which is consistent with predictions made above from blast searches, the conservation of predicted integrin-binding motifs, and the established evolutionary relationships of gnathostomes. These relationships are also observed when the C-terminal-most FNIII domain sequences are combined with the FReD sequence in the analysis (**Fig. 3B**). The 2 tenascins from the Japanese lamprey are located in an intermediate position on the tree based solely on the FReD sequences, but are placed in the tenascin-R clade when the terminal FNIII domain sequences are included (**Fig. 3B**). Note that the relative

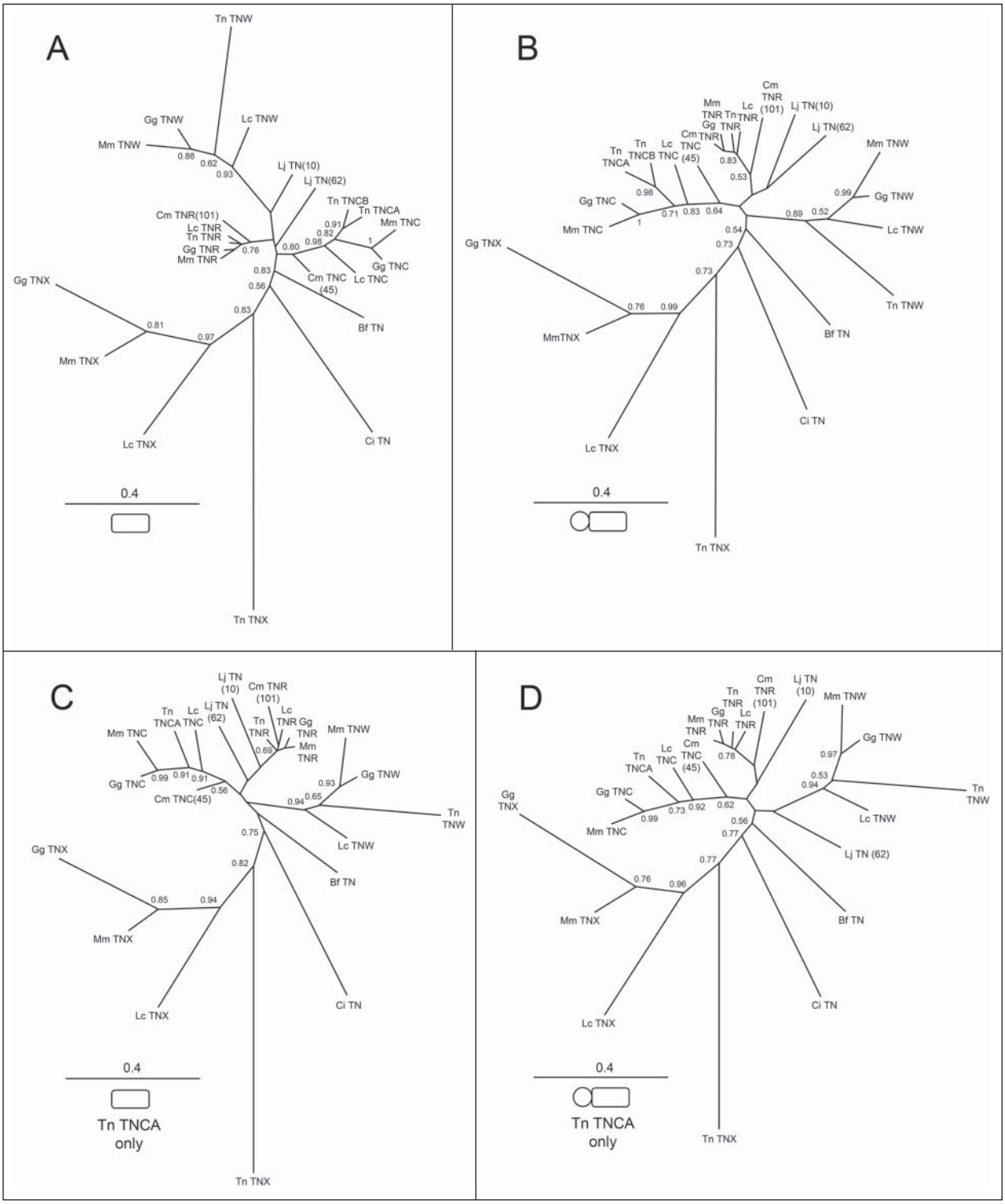

Figure 3. For figure legend, see page 27

positions of the tenascin-W and tenascin-C clades flip when the additional FNIII sequences are included in the tree construction, reflecting the low bootstrapping values associated with the branchpoints of the major tenascin family members. There are 2 tenascin-C paralogs in the green pufferfish, so additional trees were constructed after deleting the tenascin-CB paralog to balance the numbers of tenascins included in each clade. Phylogenetic trees based on the FReD amino acid sequence (**Fig. 3C**) or the FReD sequence and the C-terminal-most FNIII domain (**Fig. 3D**) generally correspond to the trees constructed using pufferfish tenascin-CA and tenascin-CB, but the placement of the Japanese lamprey tenascins varies. However, bootstrapping values are less than 0.50 for the placement of the lamprey tenascins on all 4 trees, indicating that this analysis is limited in its ability to predict the relationships between the lamprey tenascins and the tenascins of other vertebrates.

If tenascin-C lacks an RGD motif there is an RGD motif found in tenascin-W

It is well established that the RGD motif found in the F-G loop of the third FNIII domain of human tenascin-C is an integrin binding motif,[34] and that this motif is replaced by an inactive RVD in mouse tenascin-C.[36] However, mouse tenascin-W, and not human tenascin-W, has a functional RGD motif in the F-G loop of its second FNIII domain.[18] This led us to examine whether or not this pattern exists for other species that encode these 2 tenascins (**Table 2**). In actinopterygians, which have 2 tenascin-C paralogs, an RGD motif is found in some (e.g., tenascin-CB from the zebrafish *D. rerio*), but not all, third FNIII domains. RGD motifs are not found in the FNIII domains of tenascin-W from the ray-finned fishes studied here or previously.[6] Similarly, RGD motifs are missing from both tenascin-C and tenascin-W from the coelacanth *Latimeria chalmunae*. However, the pattern is commonly encountered in tetrapods. In the Western clawed frog *Xenopus tropicalis* there is an RGD motif in the F-G loop of the third FNIII domain of tenascin-W, but not in any of the FNIII domains of tenascin-C. The opposite situation (an RGD motif in the third FNIII domain of tenascin-C, but not in tenascin-W) is seen in reptiles, birds and diverse orders of mammals including carnivores, elephants, rabbits and primates. The RGD motif is missing from the third FNIII domain of tenascin-C from the superclade Cetartiodactyla, which contains cetaceans and even-toed ungulates (**Table 2** includes the Bactrian camel *Camelus ferus*, the killer whale *Orcina orca* and cattle *Bos taurus* as representatives of this group). However, in each of these species the tenascin-W has an RGD motif in the F-G loop of the second FNIII domain. In rodents the situation is more complex. Some, like the common degu *Octodon degus*, have the RGD motif in tenascin-C but not in tenascin-W. In others, like the desert jerboa *Jaculus jaculus*, the house mouse *Mus musculus*, the rat *Rattus norvegicus* and the Chinese hamster *Cricetulus griseus*, the RGD is replaced by RVD. In each of the rodents with an RVD in tenascin-C, but not in those with an RGD, there is an RGD in the F-G loop of the second FNIII domain of tenascin-W. In fact, the only group where this striking pattern is not seen is the odd-toed ungulates (Perissodactyla): in the horse *Equus ferus* and the white rhinoceros *Ceratotherium simum* both tenascin-C and tenascin-W have an RGD in the loop predicted to be exposed for integrin binding.

On the origin of fibronectin

We previously identified a fibronectin-like predicted protein in the genomic sequences of the tunicate *Ciona savignyi* that contained FNI, FNII and FNIII domains, but also contained immunoglobulin-like domains that are not seen in tetrapod fibronectins.[3] The fibronectin-like protein from *C. savignyi* also lacked the RGD motif found in the tenth FNIII domain of fibronectins from vertebrates. We now identify another fibronectin-like predicted protein with an immunoglobulin-like domain from the genomic sequence of the pelagic tunicate *Oikopleura dioica* (**Fig. 4A**). This supports the idea that fibronectin-like proteins, but not true fibronectins, are generally present in these early-diverging chordates. Each of the FNIII domains of the *O. dioica* fibronectin-like protein was analyzed by blast, and only the fourth and seventh were most similar to the FNIII domains of other fibronectins. Interestingly, many of the other FNIII domains were most similar to the FNIII domains of various vertebrate tenascins. This fibronectin-like protein does not have any of the exposed tripeptide motifs known to bind to integrins in tenascins and true fibronectins.

Figure 3 (See previous page). Phylogenetic tree analysis of tenascins. (**A**) When the coelacanth, elephant shark and Japanese lamprey tenascins are analyzed by phylogenetic tree construction based on the amino acid sequences of the fibrinogen-related domain (FReD; 229 positions in the alignment with gaps), the relationships of the coelacanth tenascins to the tenascins of other vertebrates is confirmed, and the *Callorhinchus milii* scaffold 45 tenascin clusters with tenascin-C and the scaffold 101 tenascin clusters with tenascin-R. The Japanese lamprey tenascins take an intermediate position on the tree based on the FReD, though with low bootstrapping values. (**B**) When the phylogenetic tree is based on alignment of the FReD and the C-terminal-most FNIII domain (391 positions in the alignment with gaps), most features of the phylogenetic tree based on the FReD alone are conserved, but the Japanese lamprey tenascins are placed with the tenascin-R clade and the relative positions of the tenascin-C and tenascin-W clades are flipped. (**C and D**) To balance the number of tenascins present in each clade, phylogenetic trees were also constructed without using the tenascin-CB sequence from *Tetraodon nigroviridis*. The tree based on the FReD amino acid sequence is similar to the one shown in (**B**) with both Japanese lamprey tenascins placed in the tenascin-R clade with low bootstrapping values. When the tree is constructed using the longer sequence (the FReD and the terminal FNIII domain); (**D**) one Japanese lamprey tenascin is placed in the tenascin-R clade and the other in the tenascin-W clade. The variability of the placement of the Japanese lamprey tenascins in these trees demonstrates the limited value of this analysis in predicting their relationships to the tenascins of gnathostomes. Species are abbreviated as follows: Bf, *Branchiostoma floridae* (lancelet) Ci, *Ciona intestinalis* (tunicate); Cm, *Callorhinchus milii* (elephant shark); Gg, *Gallus gallus* (chicken); Lc, *Latimeria chalumnae* (coelacanth); Lj, *Lethenteron japonicum* (Japanese lamprey); Mm, *Mus musculus* (house mouse); Tn, *Tetraodon nigroviridis* (green spotted puffer fish). Branch support higher than 0.50 is indicated. Scale = substitutions/site.

Table 2. A comparison of sequences from the third FNIII domain of tenascin-C and the second FNIII domain of tenascin-W across vertebrates

Classification		TNC FNIII-3	TNW FNIII-2
Chondrichthyes			
	Callorhincus	DTESTHSISGLEPDTEYQVSLVSNRGQMQS	n/a
Actinopterygii			
	Danio (TNCA)	SIETQYHLAELSPDTEYEVSLMARRGEMSS	EARTKHTIVGLNPGTEYQIGVQAIKGENEG
	Danio (TNCB)	ASETQYSLEDLKPDTQYRVALSSQ*RGD*VTS	
	Xiphophorus (TNCA)	SADTQYHLAGLSPDTQYEVSLTAKRGEQSS	EARTKHTIVGLYPGTEYQISVQSIKGNTKG
	Xiphophorus (TNCB)	PPDKQYNTDNLRPDTEYTVSIISRRGEATS	
Sarcopterygii			
	Latimeria	DNENQYSLGNLKPDTEYEVMLVSKRGGVRS	GENSNYLLTGLHPGTLYLITVRAIMGELEG
Amphibia			
	Xenopus	EDETQYSMNGLRPDTEYEVTLISRRREMTS	STVNNFELQDLNKGLKYTVYLLAY*RGD*RRS*
Reptilia			
	Anolis	EDESQFSIGNLKPHTEYEVTLTSR*RGD*MES	DVKSRYTITGLKPGTLYKITVISVKGEMEG
	Alligator	EDENQYSIGDLKPFTEYEVVLISR*RGD*MES	EPKSRYIITGLKPGTVYNITVIYMKDNIEG
Aves			
	Anas	EDENQYSIGNLRPHTEYEVTLISR*RGD*MES	DLKSRHIMTGLKPGTEYEVTVIPVKDGKEG
	Gallus	EDENQYSIGNLRPHTEYEVTLISR*RGD*MES	DPKSRHIMTGLKPGMEYEVTVIPVKDDIEG
Mammalia			
Carnivora			
	Mustela	QDENQYSIGNLKPDTEYEVSLISR*RGD*MSS	DPKSRYDITGLQPGTEYKITVVPMKGELEG
	Canis	HDENQYSIGSLKPDTEYEVSLISR*RGD*MSS	DPKSRYDITGLQPGTEYRITVVPMKGELEG
	Felis	HDENQYSIGNLKPDTEYEVSLISR*RGD*MSS	DPKSRYDITGLQPGTEYRITVIPMKGELEG
Proboscidea			
	Loxodonta	PDENQYSIGKLKPDTEYEVSLISR*RGD*MSS	DPKSRYDITGLHPGTEYKITVVPMKGELEG
Cetartiodactyla			
	Camelus	HEENQYSIGNLKPDTEYEVSLISHRADMSS	DPKSRYDITGLQPGTEYNITVVPM*RGD*LEG
	Orcinus	HEENQYSIGNLKPDTEYEVSLISHRADMSS	DPKSQYDITGLQPGTEYKITVVPM*RGD*LEG
	Bos	HEENQYSIGNLKPDTEYEVALISRRADMSS	DPKSRYDITGLQPGTEYKITVIPM*RGD*LEG
Perissodactyla			
	Ceratotherium	QDENQYSIGNLKPDTEYEVSLISR*RGD*MSS	DPKSRYDITGLQPGTEYKIRVVPM*RGD*LEG
	Equus	QDENQYSIGNLKPDTEYEVSLISR*RGD*MSS	DPKSRYDITGLQPGTEYKITVVPM*RGD*LEG
Lagomorpha			
	Oryctolagus	HDENQYSIGNLKPDTEYEVSLISR*RGD*MSS	EPKSRYDITGLLPGTEYKITVIPMRGELEG
Rodentia			
	Octodon	HDENQYSIGNLKPDTQYEVSLISR*RGD*MSS	EPKSRYDITGLEPGTDYKITVVPIRGELEG
	Jaculus	HEDNQYSIGNLRPDTEYEVSLISRRVDMTS	DPKSRYDITGLQPGTEYKITVVPM*RGD*LEG
	Rattus	HEDNQYSIGNLKPDTEYEVSLISRRVDMAS	DPKSRYDITGLQPGTEYKITVVPI*RGD*LEG
	Mus	HEDNQYSIGNLRPDTEYEVSLISRRVDMAS	DPKSRYDITGLQPGTEYKITVVPI*RGD*LEG
	Cricetulus	HEENQYSIGNLKPDTEYEVSLVSRRVDMAS	DPKSRYDITGLQPGTEYKITVIPI*RGD*LEG
Primates			
	Homo	EDENQYSIGNLKPDTEYEVSLISR*RGD*MSS	DPKSRYDITGLHPGTEYKITVVPMRGELEG
	Macaca	EDENQYSIGNLKPDTEYEVSLISR*RGD*MSS	DPKSRYDITGLHPGTEYKITVVPMRGTLEG

*The sequence for *Xenopus* tenascin-W comes from the third FNIII repeat, not the second FNIII repeat.

In contrast, the 2 fibronectins found in the zebrafish *Danio rerio* have the same domain architecture as the fibronectins found in tetrapods.[26] This has left an open question: can we gain any more insight into how this domain organization evolved? We identified a single predicted fibronectin protein from the genomic sequences of the coelacanth *L. chalumnae* (H3A822), the elephant shark *Callorhinchus milii* (**Fig. S3**) and the sea lamprey *P. marinus* (scaffold GL476598: 1,070,841-1,483,940). The fibronectin from the sea lamprey is striking for having an identical domain organization to the fibronectin of tetrapods, including the variable extra domains (ED) B, EDA, and IIICS (**Fig. 4A and B**). The integrin-binding RGD motif of the tenth FNIII domain is conserved in the coelacanth and elephant shark, but it is missing from the fibronectin of *P. marinus*. The *P. marinus* fibronectin does, however, have the numerous LDV motifs known to bind α4β1 integrin in tetrapod fibronectin (**Fig. 4A and B**). Thus, the "true" fibronectin appears to have co-evolved with the vertebrates and the debut of the RGD motif as a predicted integrin binding motif apparently evolved in the jawed vertebrates. Representative fibronectin domains were compared using Blastp Suite-2 sequence alignment to determine which were the most conserved (**Table 3**). The N-terminal FNI domains that are required for fibril formation are the most highly conserved in fibronectins from tetrapods and the lamprey, while the fifth and tenth FNIII domains are the least conserved, perhaps because the integrin-binding motifs represent such a small percentage of the entire amino acid sequence.

Based on the current data, the presence of fibronectin and tenascin in different classes of chordates is summarized in **Figure 5**.

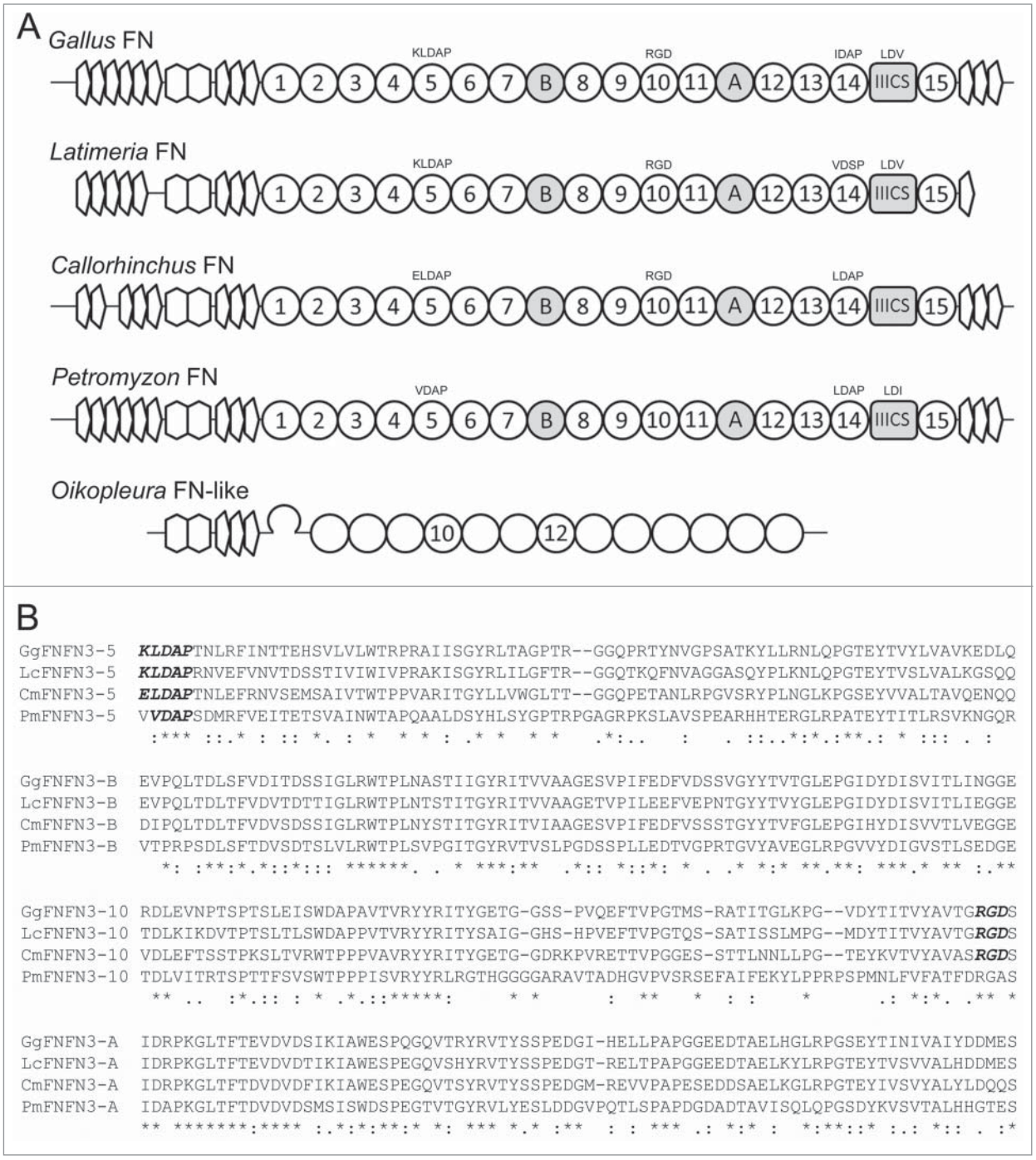

Figure 4. Analysis of novel predicted fibronectins and fibronectin-like proteins. (**A**) Fibronectin from the chicken (*Gallus gallus*), predicted fibronectins from the coelacanth (*Latimeria chalumneae*), elephant shark (*Callorhinchus milii*) and sea lamprey (*Petromyzon marinus*), and the fibronectin-like predicted protein from the pelagic tunicate *Oikopleura dioica* are shown schematically. Fibronectin-type I domains are illustrated by pentagons, fibronectin-type II domains are represented with hexagons, fibronectin-type III (FNIII) domains are represented with circles and an immunoglobulin domain is represented by a circle with an open gap, and variable domains are shaded gray. Known integrin recognition motifs are indicated above the relevant domains. (**B**) The homologous FNIII domains from the true fibronectins from different species can be aligned, illustrating the high degree of amino acid conservation and the conserved positions of integrin recognition motifs. For illustrative purposes, the fifth and tenth constant FNIII domains, as well as the variable EDB and EDA domains, are shown here.

Table 3. Comparison of amino acid sequences of representative domains from fibronectins

	Latimeria	Callorhinchus	Petromyzon
N-terminal FNI domains*			
Gallus	89% (94%)**	76% (87%)	67% (82%)
Latimeria		75% (87%)	66% (81%)
Callorhinchus			63% (75%)
FNII domains			
Gallus	80% (88%)	66% (75%)	54% (71%)
Latimeria		67% (84%)	57% (72%)
Callorhinchus			50% (68%)
FNIII5 domain			
Gallus	59% (73%)	54% (67%)	43% (64%)
Latimeria		52% (70%)	36% (59%)
Callorhinchus			32% (56%)
FNIIIB domain			
Gallus	81% (91%)	83% (91%)	53% (66%)
Latimeria		80% (93%)	56% (72%)
Callorhinchus			58% (70%)
FNIII10 domain			
Gallus	70% (77%)	58% (66%)	35% (48%)
Latimeria		58% (70%)	35% (54%)
Callorhinchus			37% (52%)
FNIIIA domain			
Gallus	81% (92%)	71% (87%)	54% (75%)
Latimeria		78% (86%)	58% (75%)
Callorhinchus			54% (74%)

*Sequence used in the analysis corresponds to the first 4 FNI domains.
**Percent amino acid identity (percent amino acid similarity). Similarity determined using blastp and the BLAST62 matrix.

Discussion

Tenascins and fibronectin play important roles in cell adhesion and migration during development and in stem cell niches in mammals. Here we have examined their representation and predicted integrin-binding capacities in early-diverging vertebrates to gain more insight into their potential contributions to the complexity of tissue organization in vertebrates. Recent phylogenomic analyses of the coelacanth *L. chalumnae* and the West African lungfish *Protopterus annectens* have concluded that while both fish belong to the extant sister group of the tetrapods, it is the latter and not the former that is most closely related to the amphibians.[31] Nevertheless, the genome of *L. chalumnae* provides important insights into the evolution of tetrapods since, for reasons yet to be determined, its protein-encoding genes are evolving very slowly.[31] Like the tetrapods, and unlike the ray-finned fishes, *L. chalumnae* has 4 tenascin genes, and these have some features (e.g.,, the additional domain that interrupts the string of FNIII domains in tenascin-X) that are more like those of the tetrapods than the tenascins from *D. rerio*.[6] However the coelacanth homologs are missing the RGD motifs found in tenascin-Cs from most tetrapods and the also the KGD motifs found in tetrapod tenascin-W, suggesting either that the coelacanth tenascins do not interact with integrins of the RGD receptor family, or that these integrins recognize other motifs in the coelacanth tenascins, or have evolved to interact with other RGD-containing ligands. As an RGD motif is present in the F-G loop of the tenth FNIII domain of coelacanth fibronectin, at least one of the coelacanth integrins is proposed to be able to interact with this motif.

Only 2 tenascins were identified in the genome of the cartilaginous elephant shark *Callorhinchus milii*: tenascin-C and tenascin-R. This is surprising because the 2 rounds of genome-wide duplications in the vertebrate lineage (2R) are estimated to have taken place before the divergence of cartilaginous fish and gnathostomes and the number of Hox gene clusters identified in elephant shark support this view.[37,38] However, the elephant shark is not an elasmobranch, the cartilaginous group that includes the sharks, skates and rays. It is a holocephalan, belonging to the subclass of Chondrichthyes that includes the rat fishes of North America, and it may have features such as the absence or selective loss of tenascins that are unique to this group and not representative of other cartilaginous fishes. It is also interesting to consider that the principal site of tenascin-W expression in birds and mammals is developing bone, and tenascin-W promotes osteogenesis *in vitro*.[20] This makes it reasonable to hypothesize that tenascin-W first appeared with the acquisition of osteogenesis in the bony fishes.

Only a single, partial, predicted tenascin was identified in the *Petromyzon marinus* genome. It is unlikely that more sequence will be found until the sea lamprey chromosomal assembly is complete, since the N-terminus of the predicted protein lies at the beginning of a genomic scaffold. The available sequence, however, reveals a very large tenascin with many similar FNIII domains. In contrast, there are at least 2 tenascins in the Japanese lamprey genome: one is likely to be the homolog of the tenascin from *P. marinus*, and the other is a smaller tenascin with at least 2 putative integrin binding motifs in its FNIII domains. The difference in the number of tenascin genes identified in the 2 species may be due to the fact that the genome of *P. marinus* was sequenced from DNA taken from adult liver prior to the remarkable finding that hundreds of millions of base pairs are deleted from somatic cell lineages during the course of lamprey embryogenesis.[39] Germ cells were used as a source of DNA for the genomic sequencing of *Lethenteron japonicum*, which may account for the presence of the second tenascin. Although initial studies concluded that the agnathan/gnathostome divergence occurred after the 2 rounds of genome duplication early in vertebrate evolution,[28,40] recent study of Hox genes in the Japanese lamprey has led to the suggestion that genome duplications may have occurred independently in the lamprey lineage.[29] This could also account for the identification of only 2 tenascins in the Japanese lamprey. The problems associated with determining the phylogenetic relationships of the lamprey tenascin sequences with each other and to other tenascins might also be accounted for by the fact that the lamprey genomes analyzed to date have unusually high GC content, which may bias the amino acid sequences of encoded proteins.

Comprehensive analyses of the FNIII domains of tenascin-C and tenascin-W reveals a remarkable pattern: if an RGD motif is missing from the F-G loop of the third FNIII domain of tenascin-C, an RGD motif is present in the F-G loop of the second or

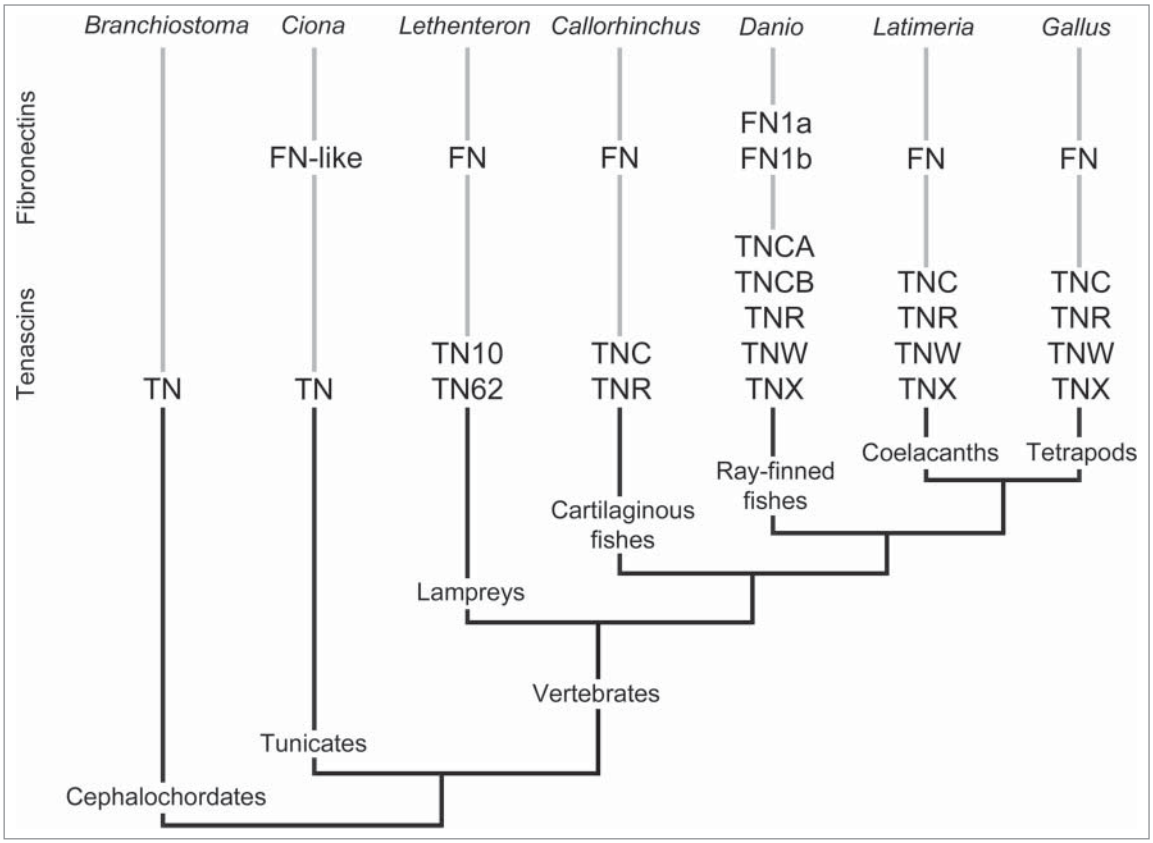

Figure 5. Current view of the representation of tenascins and fibronectin in extant chordates. The summary illustrates the probable relationships of the different groups of extant chordates and the presence or absence of fibronectin (FN), tenascin (TN), tenascin-C (TNC), tenascin-R (TNR), tenascin-W (TNW) or tenascin-X (TNX). The genera used to represent the groups are listed at the top of the tree.

third FNIII domain of tenascin-W. This suggests strongly that tenascin-C and tenascin-W might have some overlapping functions, despite the current evidence that tenascin-C/αVβ3 integrin interactions are RGD dependent, and that the RGD of tenascin-W acts through α8β1 integrin.[19] One speculation for why some species have lost the RGD sequence in their tenascin-C may be related to the observation that certain viruses attach to the cell surface by expressing RGD-containing capsid proteins. This is the case in even-toed ungulates like the cow and camel, which are prone to infection by foot-and-mouth disease virus via their αVβ3 and αVβ6 integrins,[41] and thus may be under a selection pressure to modify this receptor-ligand system. Interestingly, this same group has lost the RGD motif from tenascin-C, albeit in an ancient event prior to the evolution of cetaceans.

In the future, attention should be directed to phylogenetic analysis of the genomes of other cartilaginous fishes to determine if these animals have 4 tenascins, like the coelacanth and tetrapods, or 2 tenascins, like the elephant shark. Expressed sequence tags of the hagfish *Eptatretus burgeri*[42] encode at least one tenascin-like FReD domain, but could not be included in our phylogenetic trees because of the short sequence length. The complete hagfish genome will provide key information to include in future studies. Finally, functional studies will be needed to address the possible interactions of these newly-identified tenascins with LDV- and RGD-binding integrins.

Materials and Methods

Identification of novel predicted proteins

Predicted proteins from the genome of the coelacanth *L. chalumnae* were identified using the architecture search feature of Pfam (http://pfam.sanger.ac.uk/) and by tblastn and Delta Blast searches of coelacanth sequences at the National Center for Biotechnology Information (NCBI) (http://blast.ncbi.nlm.nih.gov/Blast.cgi) using tenascin sequences from ray-finned fishes. Predicted proteins were identified in the genome of the elephant shark *C. milii* by blast analysis at the Elephant Shark Genome Project (http://esharkgenome.imcb.a-star.edu.sg/blast/) using relevant amino acid sequences from *D. rerio* (NP_571111, XP_005171845, AAI62107, CAM13370). The blast hits were assembled by hand into predicted proteins with the same basic architecture of the proteins from bony fishes. To minimize the possibility that tenascin genes were missed in the search, searches with just the FReD sequences from tenascins were analyzed together with upstream open reading frames, none of which

revealed nearby FNIII domains except for the 2 tenascins found by the blast analysis. ESTs are not available to confirm the expression of the predicted proteins from these species. The genomic sequences of the sea lamprey *P. marinus* were searched similarly using the Ensembl Petromyzon Genome Browser (http://www.ensembl.org/Petromyzon_marinus/Info/Index) as were the genomic sequences of the Japanese lamprey *Lethenteron japonicum* (http://jlampreygenome.imcb.a-star.edu.sg/) using tenascin amino acid sequences from *D. rerio* (see above) or human (NP_002151). Predicted proteins were either complete or created by joining adjacent predicted proteins so that the resulting proteins had domain organizations similar to those found in tetrapods. The genome of the tunicate *Oikopleura dioica* (http://www.genoscope.cns.fr/externe/GenomeBrowser/Oikopleura/) was analyzed by blat searches using previously described sequences from the genus *Ciona*.[6] The sequences used for analyses can be found in **Figure S2**.

Phylogenomic analysis

Protein domains and the architecture of predicted proteins were identified using SMART (http://smart.embl-heidelberg.de/), Pfam (see above) and InterPro (http://www.ebi.ac.uk/interpro/). Some FNIII domains were aligned using ClustalW (http://www.genome.jp/tools/clustalw/) with fast pairwise alignment and other default parameters to demonstrate the conserved positions of putative integrin binding motifs. Coiled-coils were predicted using LOGICOIL (http://coiledcoils.chm.bris.ac.uk/LOGICOIL/).[32] Phylogenetic trees were created at phylogeny.fr (http://www.phylogeny.fr/version2_cgi/advanced.cgi) using MUSCLE for multiple alignment of whole predicted tenascin sequences, FReD sequences or FReD sequences combined with the most C-terminal FNIII domain. GBlocks was used for alignment curation and PhyML for tree construction. Default settings were used to construct the rooted tree with *Latimeria chalumnae* and *G. gallus* sequences, but for the unrooted trees constructed with the lamprey sequences the GBlocks settings were adjusted so as not to allow many non-conserved positions. Branch support was determined using 100 cycles of bootstrapping. The sequences used for tree construction can be found in **Figure S4**.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Supplemental Material

Supplemental data for this article can be accessed on the publisher's website.

References

1. Chiquet-Ehrismann R, Tucker RP. Tenascins and the importance of adhesion modulation. Cold Spring Harb Perspect Biol 2011; 3(5); PMID:21441591; http://dx.doi.org/10.1101/cshperspect.a004960
2. Leahy DJ, Hendrickson WA, Aukhil I, Erickson HP. Structure of a fibronectin type III domain from tenascin phased by MAD analysis of the selenomethionyl protein. Science 1992; 258:987-991; PMID:1279805; http://dx.doi.org/10.1126/science.1279805
3. Tucker RP, Chiquet-Ehrismann R. Evidence for the evolution of tenascin and fibronectin early in the chordate lineage. Int J Biochem Cell Biol 2009; 41:424-34; http://dx.doi.org/10.1016/j.biocel.2008.08.003
4. Delsuc F, Brinkmann H, Chourrout D, Philippe H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature 2006; 439:965-8; PMID:16495997; http://dx.doi.org/10.1038/nature04336
5. Schwarzbauer JE, DeSimone DW. Fibronectins, their fibrillogenesis, and in vivo functions. Cold Spring Harb Perspect Biol 2011; 3(7); PMID:21576254; http://dx.doi.org/10.1101/cshperspect.a005041
6. Tucker RP, Drabikowski K, Hess JF, Ferralli J, Chiquet-Ehrismann R, Adams JC. Phylogenetic analysis of the tenascin gene family: evidence of origin early in the chordate lineage. BMC Evol Biol 2006; 6:60; PMID:16893461; http://dx.doi.org/10.1186/1471-2148-6-60
7. Chiquet M, Fambrough DM. Chick myotendinous antigen. II. A novel extracellular glycoprotein complex consisting of large disulfide-linked subunits. J Cell Biol 1984; 98:1937-46; PMID:6202699; http://dx.doi.org/10.1083/jcb.98.6.1937
8. Grumet M, Hoffman S, Crossin KL, Edelman GM. Cytotactin, an extracellular matrix protein of neural and non-neural tissues that mediates glia-neuron interaction. Proc Natl Acad Sci U S A. 1985; 82:8075-9; PMID:2415980; http://dx.doi.org/10.1073/pnas.82.23.8075
9. Bourdon MA, Wikstrand CJ, Furthmayr H, Matthews TJ, Bigner DD. Human glioma-mesenchymal extracellular matrix antigen defined by monoclonal antibody. Cancer Res 1983; 43:2796-805; PMID:6342760;
10. Midwood KS, Hussenet T, Langlois B, Orend G. Advances in tenascin-C biology. Cell Mol Life Sci 2011; 68:3175-99; PMID:21818551; http://dx.doi.org/10.1007/s00018-011-0783-6
11. Chiquet-Ehrismann R, Orend G, Chiquet M, Tucker RP, Midwood KS. Tenascins in stem cell niches. Matrix Biol 2014; Jan 25 (e-pub ahead of print)
12. Yokosaki Y, Matsuura N, Higashiyama S, Murakami I, Obara M, Yamakido M, Shigeto N, Chen J, Sheppard D. Identification of the ligand binding site for the integrin alpha9 beta1 in the third fibronectin type III repeat of tenascin-C. J Biol Chem 1998; 273:11423-8; PMID:9565552; http://dx.doi.org/10.1074/jbc.273.19.11423
13. Yokosaki Y, Monis H, Chen J, Sheppard D. Differential effects of the integrins alpha9beta1, alphavbeta3, and alphavbeta6 on cell proliferative responses to tenascin. Roles of the beta subunit extracellular and cytoplasmic domains. J Biol Chem 1996; 271:24144-50; PMID:8798654; http://dx.doi.org/10.1074/jbc.271.39.24144
14. Huang W, Chiquet-Ehrismann R, Moyano JV, Garcia-Pardo A, Orend G. Interference of tenascin-C with syndecan-4 binding to fibronectin blocks cell adhesion and stimulates tumor cell proliferation. Cancer Res 2001; 61:8586-94; PMID:11731446
15. Orend G, Huang W, Olayioye MA, Hynes NE, Chiquet-Ehrismann R. Tenascin-C blocks cell-cycle progression of anchorage-dependent fibroblasts on fibronectin through inhibition of syndecan-4. Oncogene 2003; 22:3917-26; PMID:12813465; http://dx.doi.org/10.1038/sj.onc.1206618
16. Midwood KS, Valenick LV, Hsia HC, Schwarzbauer JE. Coregulation of fibronectin signaling and matrix contraction by tenascin-C and syndecan-4. Mol Biol Cell 2004; 15:5670-7; PMID:15483051; http://dx.doi.org/10.1091/mbc.E04-08-0759
17. Rathjen FG, Wolff JM, Chiquet-Ehrismann R. Restrictin: a chick neural extracellular matrix protein involved in cell attachment co-purifies with the cell recognition molecule F11. Development 1991; 113:151-64; PMID:1764992
18. Scherberich A, Tucker RP, Samandari E, Brown-Luedi M, Martin D, Chiquet-Ehrismann R. Murine tenascin-W: a novel mammalian tenascin expressed in kidney and at sites of bone and smooth muscle development. J Cell Sci 2004; 117:571-81; PMID:14709716; http://dx.doi.org/10.1242/jcs.00867
19. Scherberich A, Tucker RP, Degen M, Brown-Luedi M, Andres AC, Chiquet-Ehrismann R. Tenascin-W is found in malignant mammary tumors, promotes alpha8 integrin-dependent motility and requires p38MAPK activity for BMP-2 and TNF-alpha induced expression in vitro. Oncogene 2005; 24:1525-32; PMID:15592496; http://dx.doi.org/10.1038/sj.onc.1208342
20. Meloty-Kapella CV, Degen M, Chiquet-Ehrismann R, Tucker RP. Effects of tenascin-W on osteoblasts in vitro. Cell Tissue Res 2008; 334:445-55; PMID:18985388; http://dx.doi.org/10.1007/s00441-008-0715-4
21. Matsumoto K, Saga Y, Ikemura T, Sakakura T, Chiquet-Ehrismann R. The distribution of tenascin-X is distinct and often reciprocal to that of tenascin-C. J Cell Biol 1994; 125:483-93; PMID:7512972; http://dx.doi.org/10.1083/jcb.125.2.483
22. Elefteriou F, Exposito JY, Garrone R, Lethias C. Cell adhesion to tenascin-X mapping of cell adhesion sites and identification of integrin receptors. Eur J Biochem 1999; 263:840-8; PMID:10469149; http://dx.doi.org/10.1046/j.1432-1327.1999.00563.x
23. Mao JR, Taylor G, Dean WB, Wagner DR, Afzal V, Lotz JC, Rubin EM, Bristow J. Tenascin-X deficiency mimics Ehlers-Danlos syndrome in mice through alteration of collagen deposition. Nat Genet 2002; 30:421-5; PMID:11925569; http://dx.doi.org/10.1038/ng850
24. Hoegg S, Brinkmann H, Taylor JS, Meyer A. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. J Mol Evol 2004; 59:190-203; PMID:15486693; http://dx.doi.org/10.1007/s00239-004-2613-z
25. Glasauer SM, Neuhauss SC. Whole-genome duplication in teleost fishes and its evolutionary consequences.

Mol Genet Genomics. 2014; Aug 5. [Epub ahead of print]; PMID:25092473
26. Sun L, Zou Z, Collodi P, Xu F, Xu X, Zhao Q. Identification and characterization of a second fibronectin gene in zebrafish. Matrix Biol 2005; 24:69-77; PMID:15749003; http://dx.doi.org/10.1016/j.matbio.2004.12.002
27. Puech PH, Taubenberger A, Ulrich F, Krieg M, Muller DJ, Heisenberg CP. Measuring cell adhesion forces of primary gastrulating cells from zebrafish using atomic force microscopy. J Cell Sci. 2005; 118:4199-206; PMID:16155253; http://dx.doi.org/10.1242/jcs.02547
28. Smith JJ, Kuraku S, Holt C, Sauka-Spengler T, Jiang N, Campbell MS, Yandell MD, Manousaki T, Meyer A, Bloom OE, et al. Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. Nat Genet 2013; 45:415-21; PMID:23435085; http://dx.doi.org/10.1038/ng.2568
29. Mehta TK, Ravi V, Yamasaki S, Lee AP, Lian MM, Tay BH, Tohari S, Yanai S, Tay A, Brenner S, Venkatesh B. Evidence for at least six Hox clusters in the Japanese lamprey (*Lethenteron japonicum*). Proc Natl Acad Sci U S A. 2013; 110:16044-9; PMID:24043829; http://dx.doi.org/10.1073/pnas.1315760110
30. Venkatesh B, Lee AP, Ravi V, Maurya AK, Lian MM, Swann JB, Ohta Y, Flajnik MF, Sutoh Y, Kasahara M, et al. Elephant shark genome provides unique insights into gnathostome evolution. Nature 2014; 505:174-9; PMID:24402279; http://dx.doi.org/10.1038/nature12826
31. Amemiya CT, Alföldi J, Lee AP, Fan S, Philippe H, Maccallum I, Braasch I, Manousaki T, Schneider I, Rohner N, et al. The African coelacanth genome provides insights into tetrapod evolution. Nature 2013; 496:311-6; PMID:23598338; http://dx.doi.org/10.1038/nature12027
32. Zaidi AU, Kafitz KW, Greer CA, Zielinski BS. The expression of tenascin-C along the lamprey olfactory pathway during embryonic development and following axotomy-induced replacement of the olfactory receptor neurons. Brain Res Dev Brain Res. 1998; 109(2):157-68; PMID:9729356; http://dx.doi.org/10.1016/S0165-3806(98)00076-5
33. Vincent TL, Woolfson DN, Adams JC. Prediction and analysis of higher-order coiled-coils: insights from proteins of the extracellular matrix, tenascins and thrombospondins. Int J Biochem Cell Biol 2013; 45:2392-401; PMID:23891848; http://dx.doi.org/10.1016/j.biocel.2013.07.011
34. Bourdon MA, Ruoslahti E. Tenascin mediates cell attachment through an RGD-dependent receptor. J Cell Biol 1989; 108:1149-55; PMID:2466038; http://dx.doi.org/10.1083/jcb.108.3.1149
35. Campbell ID, Humphries MJ. Integrin structure, activation, and interactions. Cold Spring Harb Perspect Biol. 2011; 3(3)
36. Prieto AL, Edelman GM, Crossin KL. Multiple integrins mediate cell attachment to cytotactin/tenascin. Proc Natl Acad Sci U S A. 1993; 90:10154-8; PMID:7694284; http://dx.doi.org/10.1073/pnas.90.21.10154
37. Robinson-Rechavi M, Boussau B, Laudet V. Phylogenetic dating and characterization of gene duplications in vertebrates: the cartilaginous fish reference. Mol Biol Evol 2004; 21:580-6; PMID:14694077; http://dx.doi.org/10.1093/molbev/msh046
38. Venkatesh B, Kirkness EF, Loh YH, Halpern AL, Lee AP, Johnson J, Dandona N, Viswanathan LD, Tay A, Venter JC, Strausberg RL, Brenner S. Survey sequencing and comparative analysis of the elephant shark (*Callorhinchus milii*) genome. PLoS Biol. 2007; 5(4):e101; PMID:17407382; http://dx.doi.org/10.1371/journal.pbio.0050101
39. Smith JJ, Antonacci F, Eichler EE, Amemiya CT. Programmed loss of millions of base pairs from a vertebrate genome. Proc Natl Acad Sci U S A 2009; 106:11212-7; PMID:19561299; http://dx.doi.org/10.1073/pnas.0902358106
40. Kuraku S, Meyer A, Kuratani S. Timing of genome duplications relative to the origin of the vertebrates: Did cyclostomes diverge before or after? Mol Biol Evol 2009; 26:47-59; PMID:18842688; http://dx.doi.org/10.1093/molbev/msn222
41. Du J, Gao S, Chang H, Cong G, Lin T, Shao J, Liu Z, Liu X, Cai X. Bactrian camel (*Camelus bactrianus*) integrins alphavbeta3 and alphavbeta6 as FMDV receptors: molecular cloning, sequence analysis and comparison with other species. Vet Immunol Immunopathol. 2009; 131:190-9; PMID:19443046; http://dx.doi.org/10.1016/j.vetimm.2009.04.008
42. Takechi M, Takeuchi M, Ota KG, Nishimura O, Mochii M, Itomi K, Adachi N, Takahashi M, Fujimoto S, Tarui H, et al. Overview of the transcriptome profiles identified in hagfish, shark, and bichir: current issues arising from some nonmodel vertebrate taxa. J Exp Zool B Mol Dev Evol. 2011; 316(7):526-46; PMID:21809437; http://dx.doi.org/10.1002/jez.b.21427